STAT 201B: Probability

Feynman Liang[*]
Department of Statistics, UC Berkeley

Last updated: January 23, 2020

Contents

1 Lecture 1: A high level overview

2020-01-21

Instructor Jim Pitman, Evans 303

Text Durrett's PTE, Version Jan 2019, `https://services.math.duke.edu/~rtd/PTE/PTE5_011119.pdf`

Topics

- Mainly chapters 5–9 of Durrett's
- Material from chapter 3 neglected in 205A
- Poisson processes
- Infinitely divisible and stable laws

We will start with Markov chains; read Chapter 5 of Durrett in next few weeks.

1.1 What is a stochastic process?

Answers

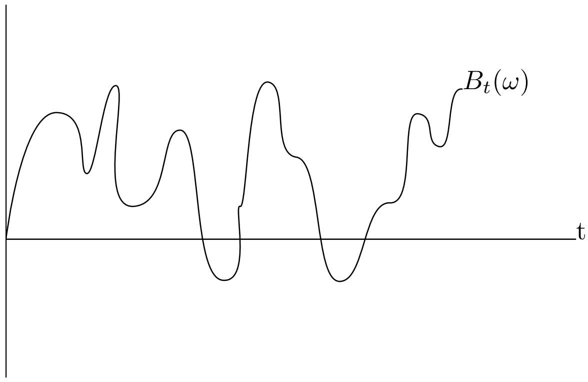
- Distribution with time
- Family of r.v.s indexed by some index set I, often $I = \text{time}, \text{space}, \text{space-time}$.
- Usually some structure on I
  - Metric space
  - $I = \{1, \ldots, n\} = [n]$ (Julia and MATLAB's choice)
    - Sometimes $I = \{0, 1, \ldots, n\}$ (Python's choice), be careful
  - Semi-group: $i, j \in I \Rightarrow i + j \in I$
    - Sometimes want a group, e.g. $I = \mathbb{Z}$
- A random function $I \to$ (some space)

[*] `feynman@berkeley.edu`

Let's unpack that last answer a bit more. Here and elsewhere, let $(\Omega, \mathcal{F}, \Pr)$ be a (background) probability space. For fixed $\omega \in \Omega$, the ***sample path*** $(X_i(\omega), i \in I)$ is a deterministic function from I to some space. To define a random function taking values in the sample paths, we must be ensure that

$$\Omega \ni \omega \mapsto (X_i(\omega), i \in I)$$

is measurable.



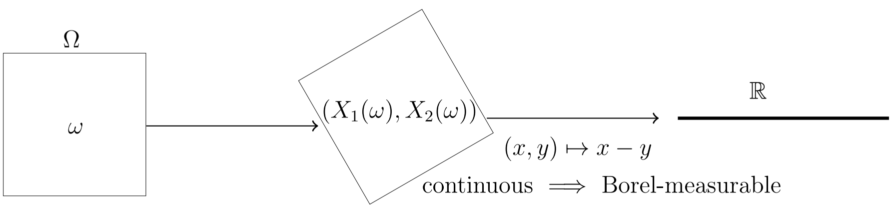
{fig:1-21-1} Figure 1: One sample path $(B_t(\omega))_{t\in I}$

Example 1

Let $I = [n]$ and suppose $(X_i, i \in I)$ are real-valued. Is $\Pr(X_i, i \in I \text{ is increasing}) = \Pr(X_1 \leq \cdots \leq X_n)$ the Pr of a measurable set?
Notice

$$\begin{aligned} \Pr(X_1 \leq \ldots \leq X_n) &= \Pr(X_{i+1} - X_i \geq 0, i \in [n-1]) \\ &= \Pr\left(\bigcap_{i\in[n-1]} \underbrace{\{X_{i+1} - X_i \geq 0\}}_{=:E_i}\right) \end{aligned}$$

Why is E_i measurable? E_i is the preimage of $[0, \infty)$ under $\omega \mapsto X_2(\omega) - X_1(\omega)$, so it suffices to show measurability of this function. But this is true because the subtraction function $(x, y) \mapsto y - x$ is measurable (because it is continuous) and composition of measurable functions remain measurable.



{fig:1-21-2} Figure 2: $\omega \mapsto X_2(\omega) - X_1(\omega)$ can be written as a composition of two measurable functions

For yet another viewpoint, consider the first half of the composition from the previous example and notice that the pair of r.v.s (X_1, X_2) induces a product σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ generated by sets of the form

$$F_1 \times F_2 = \{(x_1, x_2) : x_i \in F_i\} \tag{1}$$

for $F_i \in \mathcal{F}_i$ i.e. $X_i^{-1}(F_i) \in \mathcal{F}$. This is a very convenient generating set since (due to closure under intersection of σ-fields $\mathcal{F}_i$) it is closed under intersection and hence a π-system, enabling the use of Dynkin's π-λ theorem (taking the λ-system to be all subsets where two candidate distributions agree) to conclude the uniqueness of the distribution on the product space $\mathcal{F}_1 \otimes \mathcal{F}_2$. In this view, a stochastic process can be viewed as the unique measure on the product space $\prod_{i\in I} \operatorname{supp} X_i$.

In conclusion, we have seen that a stochastic process can be viewed as:

- A large collection of r.v.'s $(X_i, i \in I)$
- (With suitable formalism) a random function $\omega \mapsto (X_i(\omega), i \in I)$ from Ω to the product space $\prod_{i \in I} \operatorname{supp} X_i$
- The (unique) probability measure (aka ***law***) of the random function on the product space (or a suitable subset, e.g. continuous functions $\subset \{f : \mathbb{R} \to \mathbb{R}\}$)
  - This is an example of the ***push-forward*** of Pr under $\omega \mapsto (X_i(\omega), i \in I)$, which induces a probability measure on the product space
  - This allows us to forget about the background probability space $(\Omega, \mathcal{F}, \Pr)$. Instead, take Ω to be the product space, $\mathcal{F}$ the product σ-fields, and Pr the law of the push-forward measure.

Important advanced idea (French idea — Meyer & school late 1960s): You should think of the bivariate function $(\omega, i) \mapsto X(\omega, i)$ as the stochastic process.

1.2 Major classes of stochastic processes

1. IID model, $(X_i, i \in I)$ are IID; law is product of identical probability laws.
2. Independent, not necessarily identical; law is product of various probability laws. Notice there's no problem if the laws are defined with respect to different σ-fields.
3. Sums of IID or independent non-identically distributed; random walks with independent increments $S_n = \sum_{i=1}^n X_i$
4. Martingales; centering the X_i in the previous example (i.e. $\mathbb{E}X_i = 0$) makes S_n a martingale.

   Recall that a filtration is an ascending chain of σ-algebras $\mathcal{F}_n \uparrow \mathcal{F}$. For absolutely integrable random variable X, the conditional expectation (with respect to $\mathcal{F}_n$) is the (almost surely) unique random variable $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$ which integrates like X over $\mathcal{F}_n$ (important slogan), i.e.

   $$\mathbb{E}[Y_n X_n] = \mathbb{E}[Y_n X]$$

   is true for all Y_n that are $\mathcal{F}_n$-measurable (and for which the above makes sense i.e. remains integrable).

   $(M_n, \mathcal{F}_n)$ is a martingale $\Leftrightarrow$ $\mathbb{E}|M_n| < \infty$, $M_n \in \mathcal{F}_n$, and $\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = M_n$. Iterating, we have for all $i < j$ we have $\mathbb{E}_{\mathcal{F}_i} M_j = \mathbb{E}[M_j \mid \mathcal{F}_i] = M_i$

   The notation $\mathbb{E}_{\mathcal{F}_i}$ for conditional expectation is used to suggest viewing $\mathbb{E}_{\mathcal{F}_i}$ as an operator $L^1(\Omega, \mathcal{F}, \Pr) \to L^1(\Omega, \mathcal{F}, \Pr)$. In fact, $\mathbb{E}_{\mathcal{G}}$ is a partial averaging operator:

   $$\mathcal{G} \subset \mathcal{H} \Rightarrow \mathbb{E}_{\mathcal{G}}\mathbb{E}_{\mathcal{H}} = \mathbb{E}_{\mathcal{G}} = \mathbb{E}_{\mathcal{H}}\mathbb{E}_{\mathcal{G}}$$

   As a consequence, $\mathbb{E}_{\mathcal{G}}$ is a projection: $\mathbb{E}_{\mathcal{G}}\mathbb{E}_{\mathcal{G}} = \mathbb{E}_{\mathcal{G}}$.
5. Markov Chains / processes; "chain" usually implies either the state space S or time set I is discrete, whereas process often implies at least one (often both) I and S are continuous (often intervals).

   The key idea for definition of a Markov Process is conditional independence (wortth understanding in detail). Abstractly, discrete r.v.s (X, Y, Z) taking values in (S_X, S_Y, S_Z) respectively is a ***Markov triple*** iff X and Z are conditionally independent given Y, written $X \overset{Y}{\perp} Z$ means:

   (a) $\Pr[X = x, Z = z \mid Y = y] = \Pr[X = x \mid Y = y] \Pr[Z = z \mid Y = y]$ for all x, y, z such that this makes sense

   (b) $\Pr[Z = z \mid X = x, Y = y] = \Pr[Z = z \mid Y = y]$

   (c) $\Pr[X = x \mid Z = z, Y = y] = \Pr[X = x \mid Y = y]$
6. Brownian motion
7. Stationary processes (exchangeable)
8. Gaussian processes

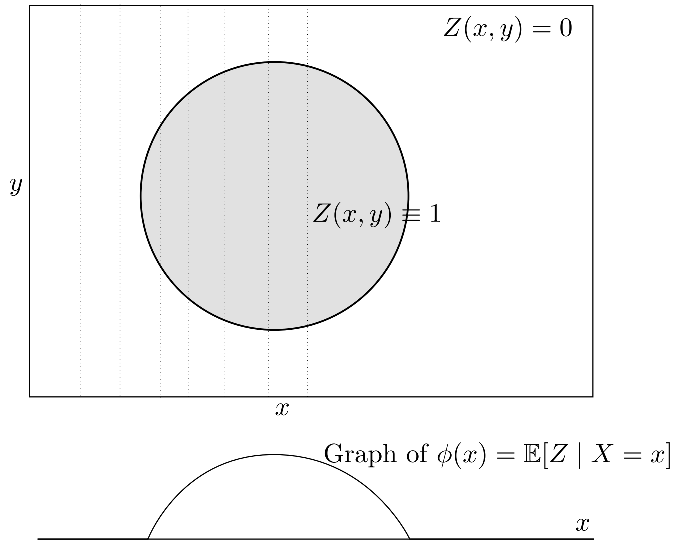


Figure 3: The conditional expectation $\phi(x)=\mathbb{E}[Z(X,Y) \mid X=x]$ can be viewed as a function of x partially averages over Y to return $\phi(x)=\mathbb{E}[Z(x,Y)]$ {fig:1-21-3}

9. Lévy processes

2 Lecture 2: Markov chains

2020-01-22

2.1 Kolmogorov consistency

A crucial tool for proving existence of sectionchastic processes is ***Kolmogorov's consistency Theorem***. Suppose we want to make a stochastic process

$$(X_i, i \in I), \quad X_i \in \mathbb{R}$$

perhaps using a large index set I.

Observe that for every $F \subset I$, $|F| < \infty$, the law of $(X_i, i \in I)$, presuming it exists on $\prod_I \mathbb{R}$, induces a joint law on $(X_i, i \in F)$.

$\Pr(\wedge_{j=1}^n X_{i_j} \le x_i)$ is known from $\prod_I \mathbb{R}$, and this is the CDF of something on $\mathbb{R}^n$. Hence, the law of $(X_i, i \in F)$ is a probability measure on $\mathbb{R}^{|F|}$. The law on $\mathbb{R}^{|F|}$ and the law on $\prod_I \mathbb{R}$ are related: ***they must be consistent***! A joint law on (X_1, X_2) implies laws for X_1 and X_2 obtained by projection (i.e. marginalization).

More generally, if $F \subset G$ then the law of $X_F := (X_i, i \in F)$ must be the projection of the law of X_G.

Definition 2

A collection of joint laws of X_F on $\mathbb{R}^{|F|}$ for each $F \subset I$ finite is ***consistent*** if it satisfies the above.

Theorem 3 (*Kolmogorov*)

For every consistent collection of finite dimensional distributions (FDDs), there exists a unique probability law on $(\mathbb{R}^I, \otimes_{i\in I}\mathcal{B})$ with these prescribed FDDs.

Remark 4. We hardly ever use it, because in nearly all interesting cases it is obvious how to construct $(X_i, i \in I)$ once you know the FDDs.

Remark 5. Unless I is countable, there is a serious issue that too few sets in $\mathbb{R}^I$ are measurable for us to care about this product construction.

Exercise 6. Show that every measurable set in $\mathbb{R}^I$ is generated by some countable set of coordinates: $B \subset \mathbb{R}^I$ in product σ-field, then $B \in (X_{i_1}, X_{i_2}, \ldots)$ for some sequence $i_1, i_2, \ldots$.

As a result, with $I = [0,1]$ the set of continuous paths is not measurable!

Example 7 (*Continuum of IID r.v.s*)

Given any consistent family of FDDs, we can construct an infinite sequence of IIDs (X_i) with any prescribed law F.
$\omega \in [0,1] \to (X_1(\omega), X_2(\omega), \ldots)$. The binary expansion of $\omega = \sum_{n=1}^{\infty} X_n(\omega) 2^{-n}$.
We can construct infintely many IID uniforms on $[0,1]$ by setting $U_1 = \sum_{i\geq 1} X_{2^i} 2^{-i}$, $U_2 = \sum_{i\geq 1} X_{3^i} 2^{-i}$, continuing with other powers of primes to maintain independence. Then use inverse CDF to construct arbitrary law.

{sub:Markov␣chains}

2.2 Markov chains

Let $(S, \mathcal{S})$ be the ***state space***, sometimes abstract. If S is countable, we will always take $\mathcal{S} = 2^S$.

Good to know what works easily for $(S, \mathcal{S})$.

Start with a general definition.

Definition 8 (*Markov transition function*)

Let $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$ be two probability spaces. P is called a ***Markov transition function / kernel*** if

$$P : S_1 \times \mathcal{S}_2 \to [0,1]$$

satisfying

1. For all $x \in S_1$, $A \mapsto P(x, A)$ is a probability measure on $(S_2, \mathcal{S}_2)$.
2. For fixed $A \in \mathcal{S}_2$, $x \mapsto P(x, A)$ is $\mathcal{S}_1$-measurable.

Usually $\mathcal{S}_2 = \sigma(\mathcal{C}_2)$ for some π-system $\mathcal{C}_2$ (closed under intersection), so we only need to check measurability of $x \mapsto P(x, A)$ for $A \in \mathcal{C}_2$. This is because (assuming 1.) $\{A : x \mapsto P(x, A) \text{ measurable}\}$ is a λ-system containing π-system $\mathcal{C}_2$, apply Dynkin π-λ.

Theorem 9

Let λ be a probability distribution on $(S_1, \mathcal{S}_1)$ and $P(\cdot, \cdot)$ a transition function for $(S_1, \mathcal{S}_1) \to (S_2, \mathcal{S}_2)$. Then $\exists!$ joint law of (X_1, X_2) on $(S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2)$ such that

$$X_1 \sim \lambda$$
$$X_2 \mid X_1 = x \sim P(x, \cdot)$$

{thm:fubini}

Theorem 10 (*Fubini*)

$$\mathbb{E}g(X_1, X_2) = \int_{S_1} \int_{S_2} g(x_1, x_2) \underbrace{\Pr(X_1 \in dx_1)}_{\lambda(dx_1)} \underbrace{\Pr(X_2 \in dx_2 \mid X_1 = x_1)}_{P(x_1, dx_2)}$$

for every product measurable g which is either bounded, non-negative, or absolutely integrable.

Exercise 11. $\mathbb{E}[g(X_1, X_2) \mid X_1] = \psi_g(X_1)$ where $\psi_g(x) = ??$

Proof. Key idea: make the joint law on $(S_1 \times S_2, \mathcal{S}_1 \times \mathcal{S}_2)$ first. Construct measure on product space, X_1 and X_2 are the coordinates.

Define

$$I(g) = \int_{S_1} \lambda(dx_1) \int_{S_2} g(x_1, x_2) P(x_1, dx_2)$$

Potential issues with $I(g)$:

- Is it defined? Try $g(x_1, x_2) = f(x_1)h(x_2)$ for $f \in m(\mathcal{S}_1)$ and $h \in m(\mathcal{S}_2)$.
- Must check the second integral is $\mathcal{S}_1$-measurable
- OK because take $f = \mathbb{1}_A$, $h = \mathbb{1}_B$.

$g \mapsto I(g)$ is linear, bounded, non-negative, monotone. What is it's measure? Take $g = \mathbb{1}_C$ for $C \in \mathcal{S}_1 \otimes \mathcal{S}_2$ and use monotonicity of the two integrals in I along with monotone convergence theorem.

Hence, $C \mapsto I(\mathbb{1}_C)$ is a probability measure on the product space satisfying the desired goals. □

Remark 12. This is only abstract measure theory. If we were in say $\mathbb{R}^2$, we could appeal to Riesz representation theorem.

How?

Flipping the transition kernel

There is no universal Bayes theorem for inverting laws (X_1), CD of $(X_2 \mid X_1) \leftrightarrow$ law (X_2), CD of $(X_1 \mid X_2)$

$$\Pr[X_2 \in A] = \int \Pr[X_1 \in dx_1] \underbrace{\Pr(X_2 \in A \mid X_1 = x)}_{P(x,A)}$$

Only if we assume densities relative to a product measure, a great ingenuity can gen flip general Markovian P to get a $\tilde{P}$.
In the discrete case, these are a matrix and vector:

$$P = (P(x, \{y\}), x \in S_1, y \in S_2) \tag{2}$$
$$\lambda = (\lambda(x), x \in S) \tag{3}$$

$$\lambda(x) = \Pr(X_1 = x) \tag{4}$$
$$\Pr(X_2 = y) = \sum_x \lambda(x) P(x, y) = (\lambda P)(y) \tag{5}$$
$$\Pr(X_1 = x \mid X_2 = y) = \frac{\lambda(x, y) P(x, y)}{(\lambda P)(y)} \tag{6}$$

for all $y : (\lambda P)(y) > 0$.

More generally, let λ be the initial distribution of X_1, P a transition kernel $X_1 \to X_2$, and $f \geq 0$ nonnegative measurable.

$$\mathbb{E} f(X_2) = \int_{S_1} \lambda(dx)(Pf)(x) = \int_{S_2} (\lambda P)(dy) f(y)$$
$$Pf(x) = \int_{S_2} P(x, dy) f(y)$$
$$(\lambda P)(A) = \int_{S_1} \lambda(dx) P(x, A)$$

Notice $\lambda \mapsto \lambda P$ and $f \mapsto Pf$ have excellent and obvious properties as operators on probability measures.

Example 13

If $f \geq 0$, then $Pf \geq 0$.
If $0 \leq f_n \uparrow f$, then $0 \leq Pf_n \uparrow Pf$.

If S_1 and S_2 are finite, then we are doing matrix algebra whree P is a $S_1 \times S_2$ stochastic matrix and λP is the action of P on a row vector λ (distributions over states, measures more generally) and Pf the action on a column vector f (observables over states, non-negative functions more generally).

Now consider $S_1 = S_2$ countable.

Definition 14

Say (X_n) is a Markov chain with initial distribution λ and stationary transition matrix P if

$$\Pr[X_0 = x_0, \dots, X_n = x_n] = \lambda(x_0) \prod_{i=1}^{n} P(x_{i-1}, x_i)$$